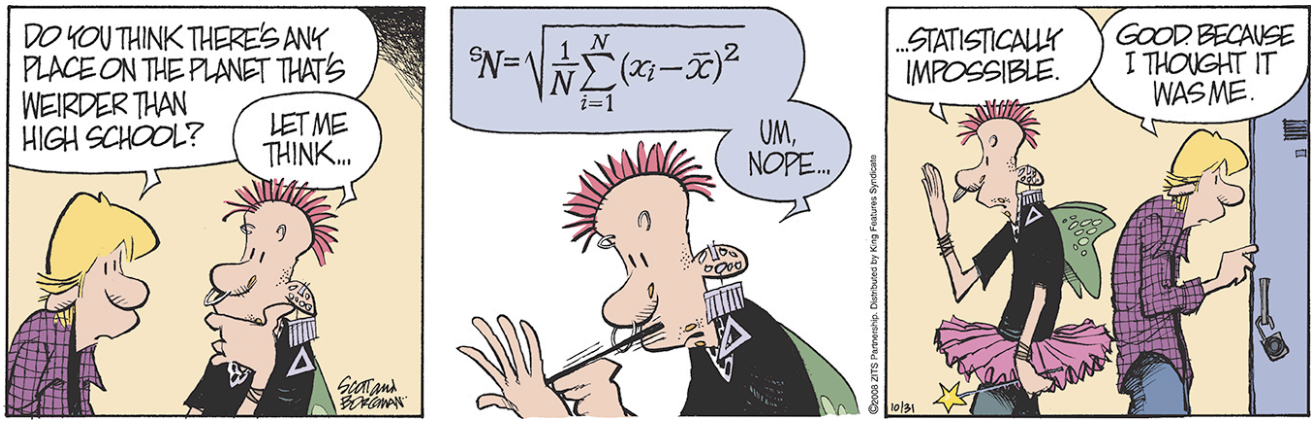
Chapter 1: Exploring Data

Key Vocabulary:

- individual
- variable
- frequency table
- relative frequency table
- distribution
- pie chart
- bar graph
- two-way table
- marginal distributions
- conditional distributions
- side-by-side bar graph
- association
- dotplot
- stemplot
- histogram
- SOCS
- outlier
- symmetric
- Σ
- x
- spread
- variability
- median
- quartiles
- Q_1, Q_3
- IQR
- five-number summary
- minimum
- maximum
- boxplot
- resistant
- standard deviation
- variance



Data Analysis: Making Sense of Data (pp.2-6)

1. *Individuals* are…

2. A *variable* is …

3. When you first meet a new data set, ask yourself:

   - Who…
   - What…
   - Why, When, Where and How…

4. Explain the difference between a *categorical* variable and a *quantitative* variable. Give an example of each.

5. Give an example of a categorical variable that has number values.

6. Define *distribution*:

7. What are the four steps to *exploring data*?
   - Begin by….
   - Study relationships…
   - Start with a …
   - Then add…

8. Answer the two questions for the *Check Your Understanding* on page 5*:*

9. Define *inference.*

1.1 Analyzing Categorical Data (pp.8-22)

1. A *frequency* table displays…

2. A *relative frequency table* displays…

3. What type of data are *pie charts* and *bar graphs* used for?

4. *Categories* in a bar graph are represented by _____ and the *bar heights* give the category _____.

5. What is a *two-way table*?

6. Define *marginal distribution*.

7. What are the two steps in examining a marginal distribution?

8. Answer the two questions for the *Check Your Understanding* on page 14.

10. What is the purpose of using a *segmented bar graph*?

9. What is a *conditional distribution*? Give an example demonstrating how to calculate one set of conditional distributions in a two-way table.

11. Answer question one for the *Check Your Understanding* on page 17.

10. What is the purpose of using a *segmented bar graph*?

12. Describe the four steps to organizing a statistical problem:

11. Answer question one for the *Check Your Understanding* on page 17.

- State…
- Plan…
- Do…
- Conclude…

13. Explain what it meant by an *association* between two variables.

12. Describe the four steps to organizing a statistical problem:

- State…
- Plan…
- Do…
- Conclude…

13. Explain what it meant by an *association* between two variables.

1.2 Analyzing Categorical Data (pp.27-42)

1. What is a *dotplot*? Draw an example.

2. When examining a distribution, you can describe the overall pattern by its

S_____ O_____ C_____ S_____

3. If a distribution is *symmetric*, what does it look like?

4. If a distribution is *skewed to the right*, what does it look like?

5. If a distribution is *skewed to the left*, what does it look like?

6. Describe and illustrate the following distributions:

    a. Unimodal

    b. Bimodal

    c. Multimodal

7. Answer questions 1-4 for the *Check Your Understanding* on page 31.

8. How are a *stemplot* and a *histogram* similar?

9. When is it beneficial to *split the stems* on a stemplot?

10. When is it best to use a *back-to-back stemplot*?

11. List the three steps involved in making a histogram.

12. Why is it advantageous to use a relative frequency histogram instead of a frequency histogram?

13. Answer questions 2-4 for the *Check Your Understanding* on page 35.

1.3 Analyzing Categorical Data (pp.50-67)

1. What is the most common *measure of center*?

2. Explain how to calculate the *mean,* $\bar{x}$.

3. What is the meaning of Σ?

4. Explain the difference between $\bar{x}$ and μ.

5. Define *resistant measure*.

6. Explain why the mean is not a resistant measure of center.

7. What is the *median* of a distribution? Explain how to find it.

8. Explain why the median is a resistant measure of center?

9. How does the shape of the distribution affect the mean and median?

10. What is the *range*?

11. Is the range a resistant measure of spread? Explain.

12. How do you find *first quartile* Q_1 and *third quartile* Q_3?

13. What is the *Interquartile Range* (IQR)?

14. Is the IQR and the quartiles a resistant measure of spread? Explain.

15. How is the IQR used to identify *outliers*?

16. What is the *five-number summary* of a distribution?

17. Explain how to use the five-number summary to make a *boxplot*.

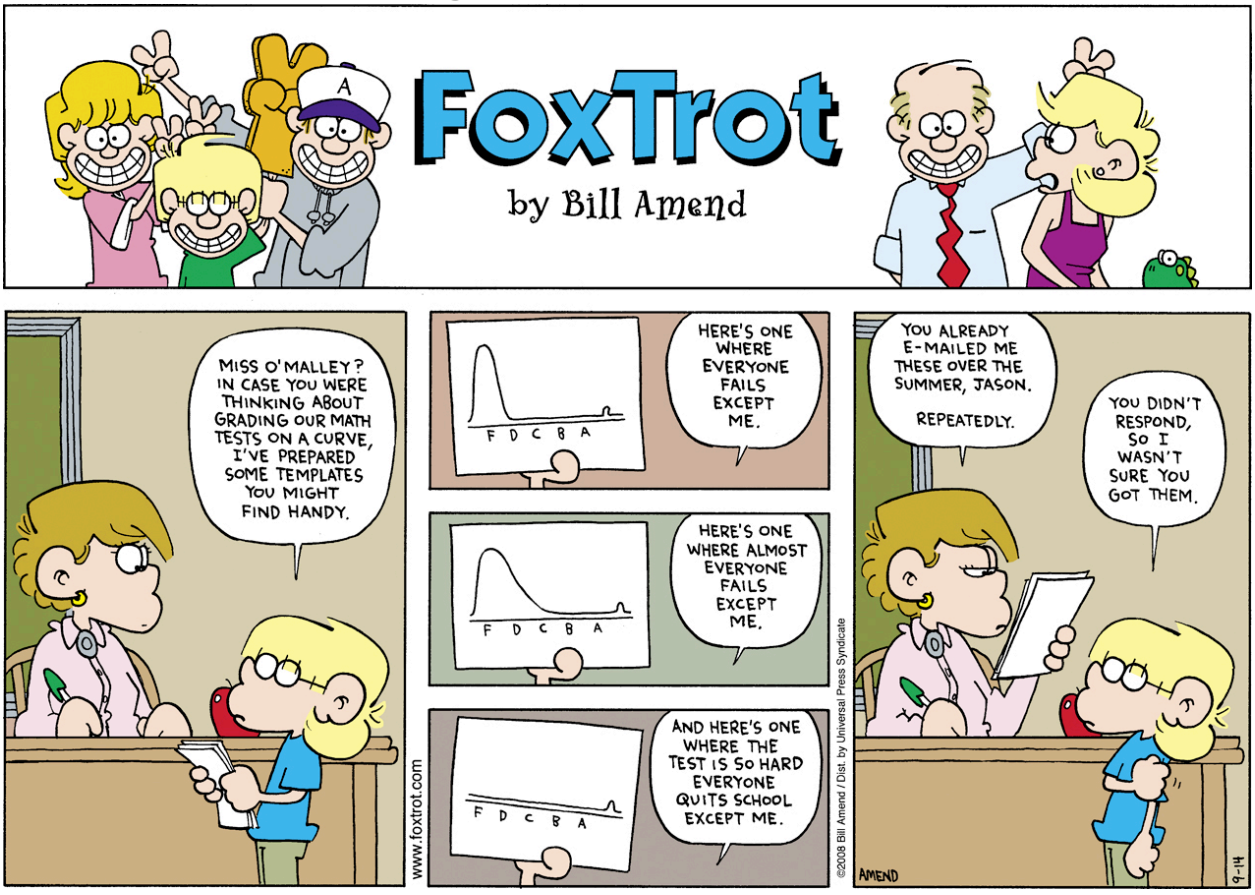
18. What does the *standard deviation* measure? How do we calculate it?

19. What is the relationship between *variance* and *standard deviation*?

20. What are the *properties* of the standard deviation as explained on page 64?

21. How should one go about choosing measures of center and spread?

Chapter 2: Modeling Distributions of Data



Key Vocabulary:

- percentiles
- cumulative relative frequency graphs
- z-scores
- transforming data
- density curves
- median of density curve
- transform data
- mean of density curve
- standard deviation of density curve
- Normal curves
- Normal distributions
- 68-95-99.7 rule
- $N(\mu,\sigma)$
- standard Normal distribution
- standard Normal table
- Normal probability plot
- μ mu
- σ sigma

2.1 Describing Location in a Distribution (pp.84-103)

1. A *percentile* is…

2. Is there a difference between the 80^{th} percentile and the top 80%? Explain.

3. Is there a difference between the 80^{th} percentile and the lower 80%? Explain.

4. Refer to the "Cumulative Relative Frequency Graphs" section on page 86 to answer the following questions:

    a. Explain how to find the *relative frequency* column.

    b. Explain how to find the *cumulative frequency* column.

    c. Explain how to find the *cumulative relative frequency* column.

5. Explain how to make a cumulative relative frequency graph.

6. What can a cumulative relative frequency graph be used to describe?

7. Answer the four questions for the *Check Your Understanding* on page 89.

8. Explain how to *standardize* a variable.

9. What information does a *z – score* provide?

10. Explain how to calculate and interpret a z- score.

11. What is the purpose of *standardizing* a variable?

12. Explain the *effects of adding or subtracting a constant* from each observation when transforming data.

13. Explain the effects of *multiplying or dividing by a constant* from each observation when transforming data.

14. Summarize the four steps for *exploring quantitative data* as outlined on page 99.

15. What is a *density curve*?

16. What does the *area* under a *density curve* represent?

17. Where is the *median* of a *density curve* located?

18. Where is the *mean* of a density curve located?

19. Answer questions 1 and 2 for the *Check Your Understanding* on page 103.

2.2 Normal Distributions (pp.110-128)

1. How would you describe the shape of a *Normal curve*? Draw two examples.

2. Explain how the mean and the standard deviation are related to the Normal curve.

3. Define *Normal distribution* and *Normal curve*.

4. What is the abbreviation for a Normal distribution with a mean μ and a standard deviation σ?

5. Explain the *68-95-99.7 Rule*. When does this rule apply?

6. Answer questions 1-3 for the *Check Your Understanding* on page 114.

7. What is the *standard Normal distribution*?

8. What information does the *standard Normal table* give?

9. How do you use the standard Normal table (Table A) to find the area under the standard Normal curve to the left of a given *z-value*? Draw a sketch.

10. How do you use Table A to find the area under the standard Normal curve to the right of a given *z-value*? Draw a sketch.

11. How do you use Table A to find the area under the standard Normal curve between two given *z-values*? Draw a sketch.

12. Summarize the steps on how to solve problems involving Normal distributions as outlined on page 120.

13. When is it appropriate to use Table A "*backwards"?*

14. Describe two methods for assessing whether or not a distribution is *approximately Normal*.

15. What is a *Normal probability plot*?

16. How do you *interpret* a Normal probability plot?

17. When is it appropriate to use the NormalCDF and Inverse Normal functions on the calculator?

Chapter 3: Describing Relationships

3.1 Scatterplots and Correlation (pp.142-156)

1. Why do we study the relationship between two quantitative variables?

2. What is the difference between a *response variable* and the *explanatory variable*?

3. How are response and explanatory variables related to *dependent* and *independent* variables?

4. When is it appropriate to use a *scatterplot* to display data?

5. A *scatterplot* shows the relationship between…

6. Which variable always appears on the horizontal axis of a scatterplot?

7. When examining a scatterplot, you can describe the overall pattern by its:

D_____ O_____ F_____ S_____

8. Explain the difference between a *positive association* and a *negative association*.

9. What is *correlation r*?

10. Answer the five questions for the *Check Your Understanding* on page 149.

11. What does correlation *measure*?

12. Explain why two variables must both be *quantitative* in order to find the *correlation* between them.

13. What is true about the relationship between two variables if the *r-value* is:

a. Near 0?

b. Near 1?

c. Near -1?

d. Exactly 1?

e. Exactly -1?

14. Is *correlation* resistant to extreme observations? Explain.

15. What do you need to know in order to *interpret* correlation?

3.2 Least-Squares Regression (pp.164-188)

1. What is a *regression line*?

2. In what way is a *regression line* a *mathematical model*?

3. What is the general form of a *regression equation*? Define each variable in the equation.

4. What is the difference between y and $\hat{y}$?

5. What is *extrapolation* and why is this dangerous?

6. Answer the four questions for the *Check Your Understanding* on page 167.

7. What is a *residual*? How do you interpret a residual?

8. What is a *least-squares regression line*?

9. What is the formula for the equation of the *least-squares regression line*?

10. The *least-squares regression line* always passes through the point ...

11. What is a *residual plot*? Sketch a graph of a residual plot.

12. If a *least-squares regression line* fits the data well, what two characteristics should the *residual plot* exhibit?

13. What is the standard deviation of the residuals? How is it interpreted?

14. How is the *coefficient of determination* defined?

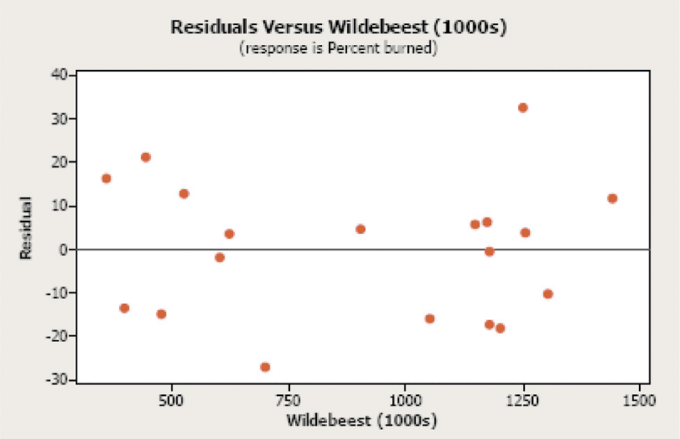
15. What is the formula for calculating the *coefficient of determination*?

16. If $r^2 = 0.95$, what can be concluded about the relationship between x and y?

_____% of the variation in (response variable) is accounted for by the regression line.

17. When reporting a regression, should r or r^2 be used describe the success of the regression? Explain.

18. Identify the *slope*, the *y intercept*, *s* and r^2 on the computer output.



Predictor	Coef	SE Coef	T	P
Constant	92.29	10.06	9.17	0.000
Wildebeest (1000s)	-0.05762	0.01035	-5.56	0.000

S = 15.9880 R-Sq = 64.6% R-Sq(adj) = 62.5%

19. What are three limitations of *correlation* and *regression*?

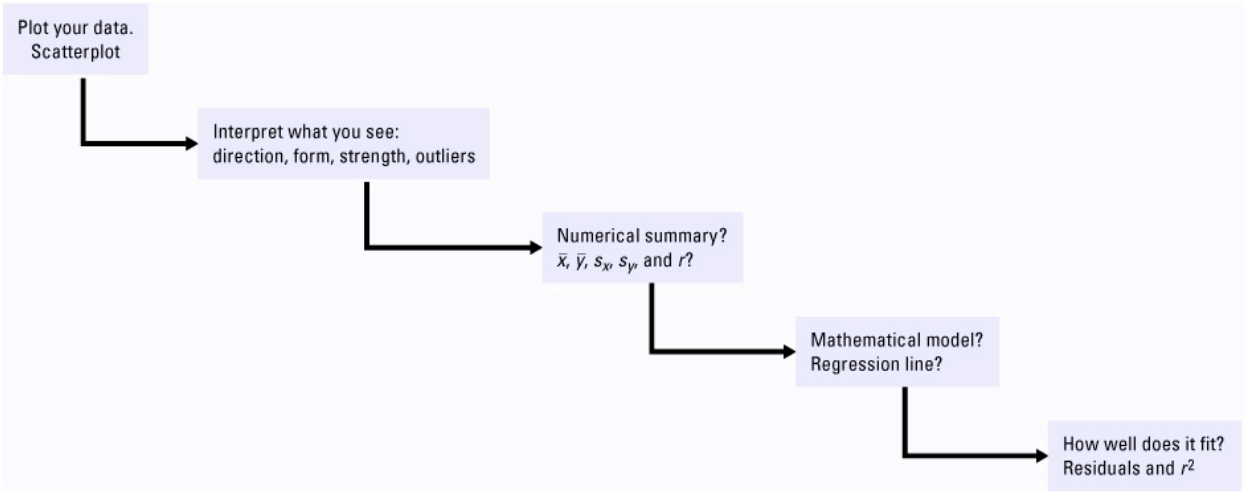
20. What is an *outlier*?

21. What is an *influential point*?

22. Under what conditions does an outlier become an *influential observation*?

23. What is a *lurking variable*?

24. Why does *association* not imply *causation*?



Chapter 4: Designing Studies

Key Vocabulary:

- sample
- population
- sample survey
- voluntary response sample
- confounded
- design
- convenience sampling
- biased
- simple random sample
- table of random digits
- probability sample
- stratified random sample
- cluster sampling
- inference
- margin of error
- strata
- undercoverage
- nonresponse
- response bias
- sampling frame
- systematic random sample
- observational study
- experimental
- confounding
- lurking variable
- experimental units
- subjects
- random assignment
- treatment
- factor
- level
- placebo effect
- single blind experiment
- control group
- completely randomized experiment
- randomized block design
- matched pair design
- statistically significant
- replication
- hidden bias
- double-blind experiment
- block design
- data ethics

4.1 Sampling and Surveys (pp.206-224)

1. Explain the difference between a *population* and a *sample*.

2. What is involved in planning a *sample survey*?

3. Why might *convenience sampling* be unreliable?

4. What is a *biased* study?

5. Why are *voluntary response samples* unreliable?

6. Define *simple random sample (SRS)*.

7. What two properties of a *table of random digits* make it a good choice for creating a simple random sample?

8. State the two steps in *choosing an SRS*:

9. What is the difference between sampling *with* replacement and sampling *without* replacement?

10. How can you account for this difference *with and without replacement* when using a table of random digits or other random number generator?

11. How do you select a *stratified random sample*?

12. What is *cluster sampling*?

13. What is *inference*?

14. What is a *margin of error*?

15. What is the benefit of a *larger* sample size?

16. A *sampling frame* is…

17. Give an example of *undercoverage* in a sample.

18. Give an example of *nonresponse bias* in a sample.

19. Give an example of *response bias* in a sample.

20. How can the wording of questions cause *bias* in a sample?

21. Answer the two questions for the *Check Your Understanding* on page 224.

4.2 Experiments (pp.231-251)

1. Explain the differences between *observational study* and *experiment.*

2. A *lurking variable* is…

3. What problems can lurking variables cause?

4. *Confounding* occurs when…

5. Answer the four questions for the *Check Your Understanding* on page 233.

6. Explain the difference between *experimental units* and *subjects*.

7. Define *treatment*.

8. By studying the TV Advertising example on page 235, identify the *factors* and *levels* in the experiment.

9. Explain why the example, *Which Works Better: Online or In-Class SAT Preparation,* is a bad experiment.

10. What is *random assignment*?

11. What is a *comparative* experimental design?

12. In a *completely randomized design*…

13. Does using chance to assign treatments in an experiment guarantee a completely randomized design? Explain.

14. What is the significance of using a *control group*?

15. The basic *principles of statistical design* experiments are:

16. Define *control*, *random assignment* and *replication* in experimental design.

17. Describe the *placebo effect*.

18. What are the differences between a *double-blind* and *single-blind* experiement?

19. Define *statistically significant*.

20. What is a *block*?

21. What is a *randomized block design*?

22. When does *randomization* take place in a block design, and how does this differ to a completely randomized design?

23. What is the goal of a *matched pairs design*?

24. When is it beneficial to use a blocked/paired design? How should we choose which variables to block for?

4.3 Using Studies Wisely (pp.261-267)

1. Name the two *types of inferences* that can be identified based on the design of a study.

2. Name the *challenges* of establishing causation.

3. What are the four criteria for *establishing causation* when we can't do an experiment?

4. Briefly describe the basics of *data ethics*.

Chapter 5: Probability: What are the chances?

Key Vocabulary:

- law of large numbers
- probability
- simulation
- two -way table
- sample space
- S = {H, T}
- tree diagram
- probability model
- replacement
- event
- P(A)
- complement A^C
- disjoint
- mutually exclusive event
- Venn diagram
- union (or)
- intersection (and)
- conditional probability
- independent events
- general multiplication rule
- general addition rule
- multiplication rule

5.1 Randomness, Probability, and Simulation (pp.282-292)

1. What is the *law of large numbers?*

2. The *probability* of any outcome…

3. How do you interpret a probability?

4. Answer the two questions for the *Check Your Understanding* on page 286.

5. What are the two *myths about randomness*? Explain.

6. Define *simulation*.

7. Name and describe the four steps in performing a simulation:

8. What are some common errors when using a table of random digits?

5.2 Probability Rules (pp.299-308)

1. In statistics, what is meant by the term *sample space*?

2. In statistics, what is meant the term *probability model*?

3. What is an *event*?

4. What is the P (A) if all outcomes in the sample space are equally likely?

5. Define the *complement* of an event. What is the complement rule?

6. Explain why the probability of any event is a number between 0 and 1.

7. What is the sum of the probabilities of all possible outcomes?

8. Describe the probability that an event does not occur?

9. When are two events considered *disjoint* or *mutually exclusive*?

10. What is the *addition rule* for mutually exclusive events?

11. What is the probability of two disjoint events?

12. Summarize the *five basic probability rules* as outlined on page 302.

13. Answer the three questions for *Check Your Understanding* on page 303.

14. When is a *two-way* table helpful?

15. In statistics, what is meant by the word *"or"?*

16. When can a *Venn diagram* be helpful?

17. What is the *general addition rule* for two events?

18. What happens if the general addition rule is used for two mutually exclusive events?

19. What does the union of two or more events mean? Illustrate on a Venn diagram.

20. What does the intersection of two or more events mean? Illustrate on a Venn diagram.

5.3 Conditional Probabilty and Independence (pp.312-327)

1. What is *conditional probablity?* What is the *notation* for conditional probabilty?

2. Answer the two questions for the *Check Your Understanding* on page 314.

3. What are *independent events*?

4. What is the *notation* used for independent events?

5. Answer the three questions for *Check Your Understanding* on page 317.

6. When is a *tree diagram* helpful?

7. State the *general multiplication rule* for any two events.

8. State the *multiplication rule* for independent events.

9. How is the *general multiplication rule* different than the *multiplication rule* for independent events?

10. Explain the difference between *mutually exclusive* and *independent*.

11. State the *formula* for calculating *conditional probabilities*.

12. How is the conditional probability formula related to the general multiplication rule?

Chapter 6: Random Variables

Key Vocabulary:

- random variable
- discrete random variable
- probability distribution
- mean of a random variable
- variance of a random variable
- probability density curve
- continuous random variable
- standard deviation
- binomial setting
- binomial random variable
- binomial distribution
- binomial coefficient
- binomial probability
- linear transformation
- normal approximation
- geometric setting
- geometric distribution
- geometric random variable
- Normal approximation
- geometric probability
- factorial
- expected value
- standard deviation
- μ_X
- μ_Y
- uniform distribution

6.1 Discrete and Continuous Random Variables (pp.341-352)

1. What is a *random variable*?

2. Define *probability distribution*.

3. What is a *discrete* random variable?

4. What are the *two requirements* for the probability distributions of discrete random variables?

5. If X is a *discrete random variable*, what information does the *probability distribution of X* give?

6. In a probability *histogram* what does the height of each bar represent (assuming the width of each bar is the same)?

7. In a probability *histogram,* what is the sum of the height of each bar?

8. What is the mean μ_X of a discrete random variable X?

9. How do you calculate the *mean of a discrete random variable*?

10. Define *expected value*. What notation is used for expected value?

11. Does the expected value of a random variable have to equal one of the possible values of the random variable? Explain.

12. Explain how to *calculate the variance and standard deviation* of a discrete random variable.

13. Explain the meaning of the standard deviation of a random variable X.

14. What is a *continuous random variable* and how is it displayed?

15. If *X* is a *continuous random variable*, how is the *probability distribution of X* described?

16. What is the area under a *probability density curve* equal to?

17. What is the difference between a *discrete random variable* and a *continuous random variable*?

18. If *X* is a *discrete random variable*, do $P(X > 2)$ and $P(X \geq 2)$ have the same value? Explain.

19. If *X* is a *continuous random variable*, do $P(X > 2)$ and $P(X \geq 2)$ have the same value? Explain.

20. How is a *Normal distribution* related to *probability distribution*?

6.2 Transforming and Combining Random Variables (pp.358-375)

1. What is the effect on a random variable of *multiplying or dividing by a constant*?

2. How does *multiplying by a constant* effect the variance?

3. What is the effect on a random variable of *adding or subtracting by a constant?*

4. Define *linear transformation*.

5. What are the effects of a *linear transformation* on the *mean* and *standard deviation*?

6. Define the *mean of the sum of random variables*.

7. What are *independent random variables*?

8. Define the *variance of the sum of independent random variables*. What types of variables does it apply to?

9. When can you *add the variances* of two random variables?

10. State the equation for the *mean of the difference* of random variables?

11. State the formula for the *variance of the difference* of random variables.

12. What happens if two independent Normal random variables are combined?

13. Suppose $\mu_X = 5$ and $\mu_Y = 10$. According to the rules for means, what is μ_{X+Y}?

14. Suppose $\mu_X = 2$. According to the rules for means, what is μ_{3+4X}?

15. Suppose $\sigma_X^2 = 2$ and $\sigma_Y^2 = 3$ and X and Y are independent random variables. According to the rules for variances, what is σ_{X+Y}^2? What is σ_{X+Y}?

16. Suppose $\sigma_X^2 = 4$. According to the rules for variances, what is σ_{3+4X}^2? What is σ_{3+4X}?

6.3 Binomial and Geometric Random Variables (pp.382-401)

1. What is a *binomial setting*?

2. Describe the *conditions* of a binomial setting.

3. What is a *binomial random variable* and what are its possible values?

4. Define the *parameters* of a binomial distribution.

5. Explain the meaning of the *binomial coefficient* and state the *formula.*

6. Explain how to *calculate binomial probabilities*.

7. What commands on the calculator are used to calculate binomial probabilities?

8. Explain how to calculate the *mean* and *standard deviation* of a *binomial random variable*.

9. When can the binomial distribution be used to sample without replacement? Explain why this is an issue.

10. What is a *geometric setting*?

11. Describe the *conditions* for a geometric setting.

12. What is a *geometric random variable* and what are its possible values?

13. Describe the *parameters* of a geometric distribution.

14. What is the *formula* for geometric probability?

15. How is the *mean* of a geometric random variable calculated?

Chapter 7: Sampling Distributions

Key Vocabulary:

- parameter
- statistic
- sampling variability
- sampling distribution
- population distribution
- biased estimator
- unbiased estimator
- bias
- variability
- variability of a statisic
- sample proportion
- mean and standard deviation of sampling distributions
- central limit theorem

7.1 What Is a Sampling Distribution? (pp.414-428)

1. Explain the difference between a *parameter* and a *statistic*?

2. What is *sampling variability*?

3. Explain the difference between μ and $\bar{x}$, and between p and $\hat{p}$?

4. What is meant by the *sampling distribution* of a statistic?

5. What is *population distribution*?

6. What is the difference between the distribution of the population, the distribution of the sample, and the sampling distribution of a sample statistic?

7. What is *sampling variability*?

8. Explain the difference between μ and $\bar{x}$, and between p and $\hat{p}$?

9. What is meant by the *sampling distribution* of a statistic?

10. What is *population distribution*?

11. What is the difference between the distribution of the population, the distribution of the sample, and the sampling distribution of a sample statistic?

12. When is a statistic considered an *unbiased estimator*?

13. What is *biased estimator*?

14. How is the size of a sample related to the *spread* of the sampling distribution?

15. The *variability of a statistic* is…

16. Explain the difference between *bias* and *variability*.

7.2 Sample Proportions (pp.432-438)

1. What is the purpose of the *sample proportion*?

2. In an SRS of size *n*, what is true about the sampling distribution of $\hat{p}$ when the sample size *n* increases?

3. In an SRS of size *n*:

a. What is the mean of the sampling distribution of $\hat{p}$?

b. What is the standard deviation of the sampling distribution of $\hat{p}$?

4. What happens to the standard deviation of $\hat{p}$ as the sample size *n* increases?

5. When does the formula $\sqrt{\frac{p(1-p)}{n}}$ apply to the standard deviation of $\hat{p}$?

6. When the sample size *n* is large, the sampling distribution of $\hat{p}$ is approximately Normal. What test can you use to determine if the sample is large enough to assume that the sampling distribution is approximately normal?

7.3 Sample Means (pp.442-452)

1. What are the mean and standard deviation of the sampling distribution of the sample mean $\bar{x}$? Describe the conditions for these formulas.

2. Explain how the behavior of the sample mean and standard deviation are similar to the sample proportion.

3. The mean and standard deviation of a population are *parameters*.
What symbols are used to represent these *parameters*?

4. The mean and standard deviation of a sample are *statistics*.
What symbols are used to represent these *statistics*?

5. The shape of the distribution of the sample mean depends on …

6. Because averages are less variable than individual outcomes, what is true about the standard deviation of the sampling distribution of $\bar{x}$?

7. What symbols are used to represent the mean and standard deviation of the sampling distribution of x ?

8. What is the mean of the sampling distribution of $\bar{x}$, if $\bar{x}$ is the mean of an SRS of size n drawn from a large population with mean μ and standard deviation σ?

9. What is the standard deviation of the sampling distribution of $\bar{x}$, if $\bar{x}$ is the mean of an SRS of size n drawn from a large population with mean μ and standard deviation σ?

10. When should you use $\frac{\sigma}{\sqrt{n}}$ to calculate the standard deviation of $\bar{x}$?

11. What is the *Central Limit Theorem*?